DOCUMENT RESUME

ED 101 685	95	IR 001 535

AUTHOR	Richards, James M., Jr.
TITLE	A Simulation Study Comparing Procedures for Assessing Individual Educational Growth. Report No. 182.
INSTITUTION	Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	JHU-CSOS-R-182
PUB DATE	Oct 74
CONTRACT	NE-C-00-3-0114
NOTE	30p.
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Computers; Educational Assessment; *Educational Research; Educational Testing; *Evaluation; Post Testing; Pretesting; Scores; Scoring Formulas; *Simulation; Teaching Techniques; Test Construction; *Testing; Testing Problems; Testing Programs; Test Interpretation; Test Results
IDENTIFIERS	ETS Growth Study

ABSTRACT

A computer simulation procedure was developed to reproduce the overall pattern of results obtained in the Educational Testing Service Growth Study. Then simulated data for seven sets of 10,000 to 15,000 cases were analyzed, and findings compared on the basis of correlations between estimated and true growth scores. Findings showed that growth was estimated more accurately by the difference between pretest and posttest scores. The measure was said to be accurate and easier to compute, and meaningful to the outsider. It was concluded that more complex assessment procedures should not be used until better results can be drawn. (Author/SK)

A SIMULATION STUDY COMPARING PROCEDURES FOR ASSESSING INDIVIDUAL EDUCATIONAL GROWTH

Contract No. NE-C-00-3-0114

Work Unit No. 2A

James M. Richards, Jr.

Report No. 182

October 1974

Published by the Center for Social Organization of Schools, supported in part as a research and development center by funds from the United States National Institute of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the Institute should be inferred.

The Johns Hopkins University

Baltimore, Maryland

STAFF

John L. Holland, Director

James M. McPartland, Assistant Director

Karl Alexander
Denise C. Daiger
David L. DeVries
Joyce L. Epstein
Ann Forthuber
Stephanie G. Freeman
Gary D. Gottfredson
Ellen Greenberger
Edward J. Harsch
Rosemary Hollick
John H. Hollifield

Ruthellen Josselson
Nancy L. Karweit
Hazel G. Kennedy
Marie Makurath
Daniel D. McConochie
Donna H. McCulloch
Edward McDill
James W. Michaels
James M. Richards
Susan L. Shackman
Julian C. Stanley

Introductory Statement

The Center for Social Organization of Schools has two primary objectives: to develop a scientific knowledge of how schools affect their students, and to use this knowledge to develop better school practices and organization.

The Center works through three programs to achieve its objectives. The Schools and Maturity program is studying the effects of school, family, and peer group experiences on the development of attitudes consistent with psychosocial maturity. The objectives are to formulate, assess, and research important educational goals other than traditional academic achievement. The School Organization program is currently concerned with authority-control structures, task structures, reward systems, and peer group processes in schools. The Careers program (formerly Careers and Curricula) bases its work upon a theory of career development. It has developed a self-administered vocational guidance device and a self-directed career program to promote vocational development and to foster satisfying curricular decisions for high school, college, and adult populations.

This report, prepared by the School Organization program, examines methods of assessing growth in achievement of individual students.

Acknowledgment

The author thanks Nancy Karweit for writing the computer programs and Truman Prevatt for carrying out analyses.

Abstract

A computer simulation procedure was developed to reproduce the overall pattern of results obtained in the ETS Growth Study. Then simulated data for seven sets of 10,000 to 15,000 cases were analyzed with several techniques for assessing growth, or change, and these techniques were compared on the basis of correlations between estimated and true growth scores. Grow h is estimated most accurately by procedures that involve the difference between the pretest and the posttest, and all estimates that involve this difference have approximately equal correlations with true growth. When one wishes to order persons on growth, there is little point in using complicated procedures to estimate growth. The simple difference between pretest and posttest scores is about as accurate as any other estimate, is much easier to compute, and should be meaningful to non-researchers. It is concluded that advocates of complex procedures should demonstrate practical, not just theoretical, advantages for their techniques before researchers should be expected to take them seriously.

Introduction

The difficulties in assessing educational growth, or psychological change, are well-known (Bloom, 1976; Campbell & Stanley, 1967; Harriott and Muse, 1973; Harris, 1963), and a variety of statistical techniques for overcoming these difficulties have been proposed (Cronbach and Furby, 1970; O'Connor, 1972). Because true growth scores typically are not known, however, it has not been possible to compare empirically the accuracy of such techniques. An obvious but largely unexploited solution to the problem of unknown true growth scores is to generate artificial data in which these scores are known.

Accordingly, in this study a computer procedure was developed to reproduce the pattern of results obtained in the ETS Growth Study (Hilton, Beaton, and Bower, 1971). Then computer-generated data were used to compare several statistical techniques for assessing growth on the basis of correlations between true and estimated growth scores. This paper treats the development of the simulation procedure and the comparison of statistical techniques as separate substudies (i.e., the Method and Results sections for these substudies are reported separately). This study is stated in the context of education, but the simulation procedures are abstract. Therefore, the results should apply to any context in which one wishes to estimate change.

Development of Simulation Procedure

Method

This study is based on the simplest additive model for growth. That is, the study is based on the following equations:

$$X_t + G_t = Y_t$$

$$X_t + E_x = X_o$$

$$Y_t + E_y = Y_o$$

using the following notation:

X_t = true pretest score

G_t = true growth score

Y_t = true posttest score

E_x = random error on pretest

X_o = observed pretest score

E_y = random error on posttest

Y_o = observed posttest score

It is important that simulated data closely resemble real data to insure that the conclusions will apply to the analysis of real data. Accordingly, this investigation aimed to reproduce the results of the ETS Growth Study (Hilton, et al., 1971) in which 9000 fifth graders in 17 communities were assessed in 1961 with both the School and College Ability Tests (SCAT) and the Sequential Tests of Educational Progress (STEP). Subject to the usual attrition in longitudinal research, these students were again assessed with STEP in 1963, 1965, and 1967. (These students were also reassessed with SCAT, but in the interests of simplicity these data are not considered.)

The present investigation treated SCAT as a measure of academic potential[1] and STEP as a measure of educational attainment. Such questions as whether SCAT is "really" a test of potential or a test of achievement, or the effects of sample attrition on Growth Study results, are largely irrelevant in this study.

The Growth Study (Hilton, et al., 1971) results include means, standard deviations, and intercorrelations for SCAT and STEP subtests. (Due to budget limitations, ETS carried out complete data analyses for males only.) For the present study these results were averaged to estimate the corresponding overall values. Then true score means, standard deviations and intercorrelations were estimated by (arbitrarily) assuming a constant reliability of .85 and applying the standard corrections for unreliability. Assuming equal reliability appears more realistic (McNemar, 1968) and leads to simpler computations than assuming that the error variances are equal.

Table 1 summarizes these computations. In this table SCAT true scores

Insert Table 1

are expressed in standard deviation units, and all STEP true scores are expressed in units of the standard deviation for Occasion 1 (i.e., the initial 1961 testing). These results show a number of features that might be expected in longitudinal studies, including an increase in average educational attainment over time, a decreasing correlation over time between

[1] The author prefers "academic potential" to "aptitude" because it reduces the implication that this is the only important dimension of talent (Richards, Holland & Lutz, 1967).

academic potential and educational attainment, and a decreasing correlation over time between earlier and later educational attainment. Humphreys (1968), for example, obtained a similar pattern of correlations for predicting early and later college GPA and Shaycoft (1967) obtained similar results, in part, from a reassessment of Project TALENT subjects.

The next step was to compute various true growth score parameters needed for the simulation from the results shown in Table 1. First the means of the growth scores for the intervals between occasions 1 and 2, 2 and 3, and 3 and 4 were computed by simple subtraction. (Throughout the rest of this paper "initial score" will be used to refer to attainment on occasion 1 and "pretest score" to attainment at the beginning of a given interval.) Then the standard deviations of the true growth scores over each interval were computed from the relationship:

$$\sigma_G^2 = \sigma_{(y-x)}^2 = \sigma_y^2 + \sigma_x^2 - 2\rho_{xy}\sigma_y\sigma_x$$

It was then possible to compute the correlations between pretest score and growth by substituting in the equation:

$$\sigma_y^2 = \sigma_x^2 + \sigma_G^2 + 2\rho_{xG}\sigma_x\sigma_G$$

and solving for ρ_{xG}. It is plausible to assume that academic potential (P) is one determinant of growth, and therefore it seemed desirable to compute the correlations between potential and growth (ρ_{pG}) over each interval. A necessary intermediate step is to compute the partial correlation between potential and growth, with current attainment held constant from the relationship:

$$\rho_{pG\cdot x} = \rho_{py\cdot x} = \frac{\rho_{py} - \rho_{px}\rho_{xy}}{\sqrt{1-\rho_{px}^2}\sqrt{1-\rho_{xy}^2}}$$

It is then possible to compute the first order correlation between potential and growth by substituting in the standard partial correlation equation:

$$\rho_{pG \cdot X} = \frac{\rho_{pG} - \rho_{pX}\,\rho_{XG}}{\sqrt{1-\rho_{pX}^2}\,\sqrt{1-\rho_{XG}^2}}$$

and solving for ρ_{pG}. It also seems possible that initial attainment continues to influence growth in the intervals 2 to 3 and 3 to 4. Accordingly, the correlations between initial score and growth (ρ_{IG}) in these intervals were computed with the same technique used to compute the correlations between potential and growth. Table 2 shows the various true growth score parameters computed by these procedures.

- - - - - - - - -

Insert Table 2

- - - - - - - - -

The usual computer procedure for generating correlated scores on two variables, A and B, generates both scores at the same time, while in the present study the A scores were given and the computer had to generate a corresponding set of B scores with the specified correlation between A and B. Accordingly, in this study B scores were generated by the following technique:

$$B = \rho_{AB} A + Z\sqrt{1-\rho_{AB}^2}$$

where both A and B have mean = 0 and standard deviation = 1, and Z is a random normal variate.[1]

The first step in using this technique to produce the simulated data was to generate a random normal variate and treat this as a given individual's true score on academic potential. Next, that individual's true score on

[1] The computer procedure for generating (pseudo) random normal variates, with mean = 0 and variance = 1, used a Fibonacci series and standardized the sum of 48 terms.

initial attainment was generated using the correlation between potential and initial attainment shown in Table 1. Then a true gain score was generated for that individual using the parameters shown in Table 2, and added to yield the true attainment score for that individual on occasion 2. Similarly, gain scores were generated and added sequentially to yield true attainment scores on occasions 3 and 4. The amount of random error implied by a reliability of .85 was added to each score, and the scores were transformed to the metric of the observed scores shown in Table 1

Simulation procedures do not always succeed in reproducing the results they are simulating. Therefore it is meaningful to evaluate such procedures by their success in reproducing such results, and in the present study to evaluate the particular assumptions used in generating data on the basis of the correspondence between Growth Study and simulated means, standard deviations, and intercorrelations. Accordingly, the parameters shown in Table 2 were used to generate simulated data under varying assumptions about how growth is determined by academic potential, initial score, and pretest score. First, three separate sets of simulated data were generated under the respective assumptions that growth is determined by each of these three characteristics alone. For example, under the assumption that growth is determined by academic potential alone, only the correlations between potential and growth shown in Table 2 were used in generating simulated data. Another three sets of simulated data were generated assuming repsectively that growth is determined by each of the three possible pairs of these variables. For each pair, the procedure was to compute the multiple regression between that pair of variables and growth. The regression weights were then used in generating the simulated data. Finally, a similar multiple regression approach was used to generate a set of simulated data under the assumption that growth is determined by a combination of all three variables. The N for each of these seven sets of simulated data was 5,000 and the sets were completely independent of each other. These N's are big enough to make questions of "significance" largely irrelevant, because almost any difference will be "significant."

Results

Table 3 compares Growth Study and simulated means and standard deviations under the various hypotheses about the determinants of growth. The correspondence between Growth Study and simulated data is reasonably close in all cases, but usually is somewhat closer when pretest score is included as one of the determinants of growth. The Growth Study and

- - - - - - - - -

Insert Table 3

- - - - - - - - -

simulated correlations are compared in Table 4. To aid in the evaluation of these comparisons, this table also applies the d^2 procedure for

- - - - - - - - -

Insert Table 4

- - - - - - - - -

measuring profile similarity (Nunnally, 1962) to the differences between each set of simulated correlations and the Growth Study correlations. (Larger d^2 values indicate greater dissimilarity.) These results indicate that the closest correspondence is obtained when potential, initial score, and pretest score are all included as determinants of growth, and that when all three are included this procedure closely reproduces Growth Study results.

These outcomes were not inevitable in the sense of being inherent in all possible growth data (e.g., simulated results determined by pretest score alone might have been equally accurate). Therefore, the results imply a number of substantive conclusions about the ETS Growth Study. The results indicate that academic potential was one of the determinants of educational growth and that the underlying relationship between potential and growth was positive. Specifically, among students with the same level of educational attainment, those students with the highest potential were likely to grow most. This trend suggests that difficult problems confront efforts to equalize educational outcomes. The continuing influence in later intervals of initial educational attainment supports Cattell's (1963) conclusion that if psychological growth is conceived of as a Markov stochastic process (Danford, Hughes, & McNee, 1960), it should not be treated as a first order process. Finally, the decrease in mean growth over successive intervals and the increasingly negative correlations between pretest score and growth suggest that the learning process tapped by this study was approaching an asymptote. In other words, the closer students were to mastering the subject matter covered by STEP, the less it was possible for them to grow.

Comparison of Statistical Procedures for Assessing Growth Method

Method

The results presented above lend credence to the use of simulated data to compare techniques for assessing growth. Accordingly, several new sets of simulated data were generated treating potential, initial score, and pretest score as the determinants of growth. These sets were generated entirely separately, and therefore can be viewed as independent replications. One set, with N = 10,000, described the situation in which all students receive the same educational treatment, or in which there are no differences among educational treatments. Six additional sets described the situation in which students are assigned to educational treatments or schools (or other social interventions) that vary in their impact on student growth. Within each of these six sets, students were assigned to 100 treatments, or schools. The number of students per school varied randomly, with mean = 150 and standard deviation = 15. Therefore, the total number of students for each of these six sets was approximately 15,000.

In three of these sets students were assigned randomly to treatments or schools, and in the other three sets students were assigned nonrandomly. When students were assigned nonrandomly, it again appeared desirable that the simulated data resemble real data as closely as possible. The most representative set of real data appeared to be the Project TALENT study of American high schools (Flanagan et al., 1962), which indicated an average correlation of approximately .54 between community per capita income and average academic potential of students. Accordingly, for each school a random normal deviate was generated and treated as the per capita income of that school's home community. Then academic potential scores for the students at that school were generated so that across schools the correlation between income and average potential was .54, and the ratio of between school variance to total variance simulated the Project TALENT ratio.

It was also assumed that community income determines school resources and that school resources in turn determine school impact. The Project TALENT data (Flanagan, et al., 1962) suggested an average correlation of approximately .25 between community income and those school resources presumed to facilitate student growth, and accordingly this correlation was used in deriving the simulated data. No data are available, however to estimate the correlation between school resources and school impact. Therefore, simulated data were computed under three different assumptions about this relationship. Specifically, it was assumed that school resources account for 5%, 20%, or 80% of the variance in school impact (corresponding to correlations of .2236, .4472, or .8944). These three assumed relationships by the two kinds of assignment (random or nonrandom) defined the six sets of simulated data.

Within each set, it was assumed that school impact is normally distributed, and that average growth scores are the same as those shown in Table 2 for a school with average impact and 10% higher for a school one standard deviation above the mean on impact. The magnitude of this standard deviation for school impact is the approximate value at which, given the values shown in Table 2, two schools one standard deviation apart on impact (with N's = 150) will differ significantly at the .05 level. In computing the individual growth scores, the averages shown in Table 2 were adjusted in accordance with school impact and no other changes were made.

The amount of growth over various intervals for individuals was estimated from the observed scores within each set of simulated data by nine different techniques. The first eight of these were taken from the article by Cronbach and Furby (1970) and use the equations for "unlinked" scores presented in that article. The ninth estimate is a standard adjustment of outcome for initial academic potential. The multiple correlational estimates of growth outlined by Cronbach and Furby were not considered because McNemar (1958) has shown analytically that such estimates are not notably more accurate than simple residual gain.

These nine estimates include:

1. Posttest score.
2. Raw gain. This gain score is the simple difference between the posttest and the pretest.
3. Gain adjusted for pretest error. This gain score is the difference between the posttest and estimated true score on the pretest.
4. Gain adjusted for pre- and posttest error. Obviously, this measure of growth is the difference between estimated true posttest and pretest scores.
5. Lord (1956, 1958) procedure. This technique provides an estimate of the difference between true posttest and pretest scores (which is not the same as the difference between estimated true pre- and posttest scores.)
6. Raw residual gain. This growth score is the difference between the posttest and predicted score on the posttest, using the pretest as predictor. Thus, this technique resembles analysis of covariance, with the pretest treated as the covariate.

7. Estimated true residual gain. This technique provides the estimated difference between true score on the posttest and the posttest score predicted from true scores on the pretest.

8. Tucker-Damarin-Messick (1966) "basefree" procedure for measuring change. This technique was designed more for correlational studies than for providing interpretable estimates of individual gain (Cronbach and Furby, 1970).

9. Posttest score adjusted for initial academic potential. This procedure is identical with the raw residual gain except the predicted posttest score is based on academic potential rather than on the pretest.

The most reasonable basis for evaluating the accuracy of these estimates is their correlations with true growth scores. The mean and standard deviation of the growth scores can be estimated more efficiently by the kind of direct procedure used in deriving Table 2. Moreover, the formula for each of the complex estimates of growth involves adding a constant to equate the mean of that estimate with mean raw growth (Cronbach and Furby, 1970), and a similar simple transformation could equate the standard deviations. Accordingly, within each set of simulated data correlations were computed between true and estimated gain scores.

Results

Table 5 shows the correlations between the nine estimated growth scores

- - - - - - - -

Insert Table 5

- - - - - - - -

and the corresponding true growth scores when all students are assigned to the same educational treatment. Table 6 shows the correlations when students are assigned to educational treatments that vary in their impact on students. Two estimates of growth are eliminated from Table 6 because of their high redundancy with other estimates, especially the Lord procedure.

\- - - - - - - -

Insert Table 6

\- - - - - - - -

These results are highly consistent and stable (in part an indication of the utility of large N's), and seem quite unequivocal. All nine estimates of growth are correlated at least moderately with true growth, and the magnitude of the correlation increases as the growth interval increases. Growth is estimated most accurately by procedures that involve the difference between the pretest and the posttest, and all procedures that involve this difference have approximately equal correlations with true growth. This is the case even when students are assigned nonrandomly to educational programs with varying impact.

Discussion

The results of this study clearly indicate the usefulness of simulation procedures in studies of the methodology for assessing growth. Analytic treatments of these issues have emphasized the theoretical advantages of various estimates of growth, and have paid little attention to the practical differences between estimates. In particular, most analytic treatments have emphasized the supposed disadvantages of raw gain scores, and have implied that quite different results would be obtained with some other estimate of growth. Cronbach and Furby (1970) are virtually alone in questioning (on different grounds than those raised in this paper) the need for the more complex procedures and certainly examination of their equations suggests the practical differences between estimates must be small. The simulation procedures used in this study show that these differences are very small indeed.

Simulation procedures also have important advantages over "real" longitudinal studies. To get longitudinal data for 10,000 to 15,000 subjects over a six-year period would require a massive investment of staff and money and would involve massive problems in keeping track of students, test scoring, tape merging, and the like. Moreover, at the end of this enterprise one usually still would not know the true growth scores for individuals. Generating longitudinal data for each set of 10,000 to 15,000 subjects for the present study required less than ten minutes on an obsolescent computer (IBM 7094), the true growth scores for individuals were known, and the procedures could easily be extended to the investigation of such questions as the effects of sample attrition, the most appropriate way to aggregate data for groups of subjects, etc. There appears to be little doubt, therefore, that simulation techniques should provide the procedure of choice in most empirical investigations of longitudinal methodology.

More important, the implications of the results for choice of a procedure to assess growth, or change, seem clear. When one wishes to order persons on growth, there is little point in using complicated estimates of growth. The simple difference between the pretest and the posttest is about as accurate as any other estimate, is much easier to compute, and should be immediately meaningful to nonresearchers. The explanation for this finding probably is relatively simple. When fallible pretest and posttest scores are the only information one has with which to estimate individual growth, it is possible to estimate group true score parameters by the kind of direct technique used in deriving table 2, but no amount of statistical legerdemain with variances, reliabilities, and intercorrelations will increase the amount of information in or remove the fallibility of the individual scores (i.e., there is no way of knowing whether an individual score involves positive or negative error, nor the magnitude of that error). Therefore, such legerdemain is unlikely to produce a set of individual growth estimates notably more accurate than one based directly on the fallible scores. Indeed, the derivation of ever more esoteric formulas for such legerdemain may be a negative contribution because it may intimidate investigators who could use simple techniques with equal accuracy and legitimacy. Therefore, advocates of complex procedures should demonstrate practical, not just theoretical, advantages for their techniques before researchers can be expected to take them seriously.

References

Bloom, B.S. Stability and Change in Human Characteristics. New York: Wiley, 1964.

Campbell, D.T. and Stanley, J.C. Experimental and Quari-Experimental Designs for Research. Chicago: R. McNally, 1967.

Cattell, R.B. The structuring of change by p-techniques and incremental r-technique. In C.W. Harris, Ed. Problems in Measuring Change. Madison, Wisconsin: University of Wisconsin Press, 1963.

Cronbach, L.J. and Furby, L. How should we measure change--or should we? Psychological Bulletin, 1970, 74, 68-80.

Danford, M.B.; Hughes, H.; and McNee, R.C. The analysis of repeated measures experiments. Biometrics, 1960, 16, 547-565.

Flanagan, J.C.; Dailey, J.T.; Shaycoft, M.F.; Orr, D.B.; & Goldberg, I. Studies of the American High School. Project TALENT Monograph No. 2. Pittsburgh: University of Pittsburgh, 1962.

Harris, C.W. (Ed.) Problems in Measuring Change. Madison, Wisconsin: University of Wisconsin Press, 1963.

Herriott, E.H. and Muse, D.N. Methodological issues in the study of school effects. In F.R. Kerlinger, Ed. Review of Research in Education. Itasca, Illinois: Peacock, 1973.

Hilton, T.L.; Beaton, A.E.; and Bower, C.P. Stability and Instability in Academic Growth--A Compilation of Longitudinal Data. Princeton, New Jersey: Educational Testing Service, 1971.

Humphreys, L.G. The fleeting nature of the prediction of college academic success. Journal of Educational Psychology, 1968, 59, 375-380.

Lord, F. M. The measurement of growth. _Educational and Psychological Measurement_, 1956, _16_, 421-437.

Lord, F. M. Further problems in the measurement of growth. _Educational and Psychological Measurement_, 1958, _18_, 437-451.

McNemar, Q. On growth measurement. _Educational and Psychological Measurement_, 1958, _18_, 47-55.

Nunnally, J. C. The analysis of profile data. _Psychological Bulletin_, 1962, _59_, 311-319.

O'Connor, E. F. Extending classical test theory to the measurement of change. _Review of Educational Research_, 1972, _42_, 73-98.

Richards, J. M., Jr.; Holland, J. L.; & Lutz, S. W. The prediction of student accomplishment in college. _Journal of Educational Psychology_ 1967, _58_, 343-355.

Shaycoft, M. F. _The High School Years: Growth in Cognitive Skills_. Pittsburgh: American Institutes for Research, 1967.

Tucker, L. R.; Damarin, F.; & Messick, S. A base-free measure of change. _Psychometrika_, 1966, _31_, 457-473.

Table 1

Summary of Results of ETS Growth Study

		Academic Potential	Occasion 1	2	3	4	Mean	S.D.
Academic Potential (SCAT)		-	.75	.70	.66	.63	253.28	10.28
Educational Attainment (STEP) on Occasion:								
1	(1961)	.88	--	.75	.72	.67	255.17	13.16
2	(1963)	.82	.88	--	.76	.70	266.48	14.65
3	(1965)	.78	.85	.89	--	.74	276.72	15.63
4	(1967)	.74	.79	.82	.87	--	284.10	16.16
Mean		0.000	0.000	0.866	1.687	2.338		
S.D.		1.000	1.000	1.116	1.194	1.224		

Note: Average of observed score values obtained in ETS study are shown above the diagonal and estimated true score values computed for this study are shown below the diagonal. True scores are expressed in standard deviation units.

Table 2

True Growth Score Parameters

Computed from ERS Growth Study Data

	Growth Between Occasions		
Parameter	1 and 2	2 and 3	3 and 4
Growth Mean	.866	.821	.651
Growth S.D.	.530	.547	.617
$\rho_{X\theta}$	-.034	-.098	-.209
$\rho_{p\theta \cdot X}$	.20	.20	.20
$\rho_{I\theta \cdot X}$	--	.31	.19
$\rho_{p\theta}$	.065	.034	-.041
$\rho_{I\theta}$	--	.060	-.078

Note: For growth between occasions 1 and 2, $\rho_{I\theta}$ is identical with $\rho_{X\theta}$.

Table 3

Comparison of ETS Growth Study and Corresponding Simulated Observed Score Means and Standard Deviations Under Various Hypotheses About Determinants of Growth

	Academic Potential		Occasions							
			1		2		3		4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ETS Growth Study	253.28	10.28	255.17	13.16	266.48	14.65	276.72	15.63	284.10	16.16
Growth Determined by: (N's = 5,000)										
Potential	253.19	10.31	255.27	13.34	266.10	15.55	276.36	17.15	284.17	18.70
Initial Score	253.14	10.19	255.31	13.28	266.12	14.74	276.27	16.60	284.14	17.81
Pretest Score	253.34	10.48	255.40	13.54	266.02	14.90	276.04	15.89	284.35	16.57
Potential & Initial Score	253.41	10.23	255.47	13.39	266.37	14.85	276.41	15.75	284.59	16.33
Potential & Pretest Score	253.42	10.24	255.66	13.26	266.34	14.69	276.53	17.70	284.67	17.37
Initial & Pretest Score	253.38	10.43	255.54	13.43	266.21	14.94	276.45	16.00	284.76	16.12
Potential, Initial & Pretest Score	253.31	10.21	255.49	13.39	266.10	14.98	276.19	15.88	284.48	16.32

Note: "Initial Score" refers to attainment on Occasion 1 and "Pretest Score" to attainment at the beginning of a given interval.

Table 4

Comparison of ETS Growth Study and Corresponding Simulated Observed Score Correlations
Under Various Hypotheses about Determinants of Growth

		Simulated Data (N's = 5,000 each)						
		Growth Determined By:						
Potential vs.	Growth Study	Potential	Initial Score	Pretest Score	Potential & Initial Score	Potential & Pretest Score	Initial & Pretest Score	Potential, Initial & Pretest Score
Occasion 1	.75	.74	.74	.75	.74	.75	.75	.74
Occasion 2	.70	.68	.66	.66	.69	.70	.66	.68
Occasion 3	.66	.62	.60	.58	.65	.67	.65	.66
Occasion 4	.63	.56	.53	.51	.52	.63	.58	.64
Occasions:								
1 vs. 2	.75	.76	.75	.75	.75	.74	.75	.75
1 vs. 3	.72	.70	.69	.67	.71	.68	.72	.72
1 vs. 4	.67	.63	.60	.60	.58	.63	.66	.68
2 vs. 3	.76	.79	.77	.76	.78	.76	.76	.75
2 vs. 4	.70	.71	.69	.65	.66	.68	.67	.69
3 vs. 4	.74	.79	.77	.74	.76	.75	.73	.74
	$\sqrt{\Sigma d^2}$	.11	.15	.18	.15	.06	.07	.03

Table 5

Correlations Between True Gain and Gain Estimated from Observed Scores by Various Procedures When All Students are in Same Educational Program

(N = 10,000)

Estimation Procedure	Growth During Interval 1 to 2	1 to 3	1 to 4	2 to 3	2 to 4	3 to 4
1. Posttest Score	.40	.50	.52	.34	.41	.26
2. Raw Gain	.64	.68	.74	.61	.73	.65
3. Gain Adjusted for Pretest Error	.65	.70	.74	.62	.73	.64
4. Gain Adjusted for Pre- and Posttest Error	.64	.68	.74	.61	.73	.65
5. Lord Procedure	.65	.70	.74	.62	.73	.65
6. Raw Residual Gain	.65	.70	.74	.62	.71	.62
7. Estimated True Residual Gain	.65	.70	.74	.62	.71	.62
8. Tucker-Damarin-Messick Procedure	.64	.68	.74	.62	.73	.65
9. Posttest Score Adjusted for Initial Academic Potential	.51	.59	.63	.42	.52	.36

Table 6

Correlations between True Gain and Gain Estimated from Observed Scores by Various Procedures When Students are Assigned to Educational Programs with Varying Impacts

(N's = Approximately 15,000)

Estimate of Growth	Correlations Between Resources and Impact	Growth in the Interval											
		1 to 2		1 to 3		1 to 4		2 to 3		2 to 4		3 to 4	
		R	N	R	N	R	N	R	N	R	N	R	N
Posttest Score	.2236	.40	.42	.51	.51	.53	.55	.35	.35	.40	.41	.26	.28
	.4472	.42	.44	.53	.53	.54	.55	.36	.35	.41	.40	.26	.27
	.8944	.42	.42	.52	.52	.54	.54	.36	.35	.40	.41	.27	.27
Raw Gain	.2236	.65	.65	.70	.70	.74	.75	.63	.63	.73	.73	.65	.66
	.4472	.65	.65	.71	.70	.74	.74	.64	.63	.73	.72	.65	.65
	.8944	.65	.64	.71	.70	.74	.74	.63	.63	.72	.73	.66	.65
Lord Procedure	.2236	.66	.66	.72	.71	.75	.76	.63	.63	.73	.73	.65	.66
	.4472	.66	.66	.72	.72	.75	.75	.64	.64	.73	.73	.65	.65
	.8944	.66	.66	.72	.71	.75	.75	.63	.63	.72	.73	.66	.65
Raw Residual Gain	.2236	.65	.66	.72	.71	.75	.76	.63	.63	.71	.71	.62	.63
	.4472	.66	.66	.72	.72	.75	.75	.64	.63	.71	.71	.62	.62
	.8944	.66	.66	.72	.71	.75	.75	.63	.62	.71	.71	.62	.62
Estimated True Residual Gain	.2236	.65	.66	.72	.71	.75	.76	.63	.63	.71	.71	.61	.63
	.4472	.66	.66	.72	.72	.75	.75	.64	.63	.71	.71	.62	.62
	.8944	.66	.66	.72	.71	.75	.75	.63	.62	.71	.71	.62	.61
Tucker-Damarin-Messick Procedure	.2236	.65	.65	.70	.70	.75	.75	.63	.63	.75	.73	.65	.65
	.4472	.65	.65	.70	.70	.74	.74	.64	.63	.73	.73	.65	.65
	.8944	.65	.65	.71	.69	.74	.74	.63	.63	.72	.73	.65	.64
Posttest Score Adjusted for Initial Academic Potential	.2236	.52	.53	.62	.62	.65	.66	.44	.44	.51	.52	.35	.36
	.4472	.52	.52	.62	.62	.64	.65	.44	.44	.51	.52	.35	.36
	.8944	.51	.52	.62	.62	.65	.65	.45	.43	.52	.51	.36	.36

NOTE: Columns headed R show correlations when students are assigned randomly to educational treatments and columns headed N show correlations when students are assigned nonrandomly.